



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ciencias Matemáticas

Escuela Profesional de Estadística

**Comparación de árboles de clasificación CART y redes
neuronales perceptrón multicapa para la clasificación
de estudiantes de secundaria que consumen o no
alcohol**

TESINA

Para optar el Título Profesional de Licenciada en Estadística

AUTOR

Ingrid VILLANUEVA GALVÁN

ASESOR

Geraldine Judith VIGO CHACÓN

Lima, Perú

2017



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Villanueva, I. (2017). *Comparación de árboles de clasificación CART y redes neuronales perceptrón multicapa para la clasificación de estudiantes de secundaria que consumen o no alcohol*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas, Escuela Profesional de Estadística]. Repositorio institucional Cybertesis UNMSM.

U-R.
62-P.

UNIVERSIDAD NACIONAL MAYOR DE SANMARCOS
(Universidad del Perú, DECANA DE AMERICA)

FACULTAD DE CIENCIAS MATEMATICAS

PROGRAMA DE ACTUALIZACIÓN PARA LA TITULACIÓN PROFESIONAL 2017- I
MODALIDAD EXAMEN DE SUFICIENCIA PROFESIONAL

ESCUELA PROFESIONAL DE ESTADÍSTICA

ACTA DE EXPOSICIÓN DE TESINA

En la Ciudad Universitaria, Facultad de Ciencias Matemáticas, siendo las 11:30 horas, del día 26 de agosto del 2017, se reunieron las docentes designadas como miembros del Jurado Evaluador:

- | | |
|---------------------------------|------------|
| - Lic. Grabiela Montes Quintana | Presidenta |
| - Lic. Geraldine Vigo Chacón | Miembro |

Para la exposición de Tesina titulada: «COMPARACIÓN DE ÁRBOLES DE CLASIFICACIÓN CART Y REDES NEURONALES PERCEPTRON MULTICAPA PARA LA CLASIFICACIÓN DE ESTUDIANTES DE SECUNDARIA QUE CONSUMEN O NO ALCOHOL», presentada por la Bachiller **Ingrid Villanueva Galván**.

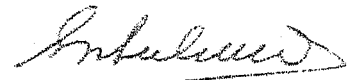
Luego de la exposición de la tesina, los Miembros del Jurado hicieron las preguntas correspondientes, a las cuales la Bachiller **Ingrid Villanueva Galván**, respondió con acierto y solvencia, demostrando pleno conocimiento del tema.

Hecha la evaluación correspondiente, según tabla adjunta, la Bachiller **Ingrid Villanueva Galván** mereció la aprobación obteniendo como calificativo promedio y la nota de diecisiete (17) (letras y números).

A continuación los miembros del Jurado, dan manifiesto que la Bachiller **Ingrid Villanueva Galván** APROBÓ la exposición de la tesina.

Siendo las 12:00 horas, se levantó la sesión, firmando para constancia la presente acta en dos (2) copias originales.


Lic. Geraldine Vigo Chacón
MIEMBRO


Lic. Grabiela Montes Quintana
PRESIDENTA

RESUMEN

El propósito del presente estudio es mostrar que las técnicas de minería de datos pueden ser una alternativa para la clasificación de los estudiantes de secundaria que consumen o no alcohol. Las técnicas que se usarán son Árboles de clasificación CART y Redes Neuronales de tipo Perceptrón Multicapa.

Se aplicarán las dos técnicas en una población de 1044 estudiantes de secundaria, y esta, será dividida en dos muestras, el 70% para el entrenamiento y 30% para la validación.

Para la comparación de los modelos que mejor discrimina a los estudiantes de secundaria que consumen o no alcohol, se utilizará la Sensibilidad y el Coeficiente de Gini. Adicionalmente, se hará la comparación aplicando la validación cruzada en ambos modelos.

Palabras clave: Redes Neuronales Perceptrón Multicapa, Arboles clasificación CART, Gini, Sensibilidad.

ABSTRACT

The purpose of this study is to demonstrate that data mining techniques are an alternative for classification of high school students who drink and those who do not drink alcohol. The techniques to be used are Classification and Regression Trees (CART) and Multi-layer Perceptron Neural Networks.

The two techniques will be applied in a population of 1044 secondary students and it will be divided into two samples, 70% for training and 30% for validation. Sensitivity and Gini coefficient will be used to determine the model that best discriminates the high school students who drink and those who do not drink alcohol. Additionally, comparison will be performed using cross validation in both models

Key words: Neural networks Multi-Layer Perceptron, CART classification tree, Gini, Sensitivity.